

# CHAPTER 4

## ANALYSIS AND DESIGN

### 4.1 Analysis

Analysis is the process of solving a complex problem into small parts so that it can be more easily understood. In this project the analysis process is divided into five parts. the first is the data collection process. Then the data simplification process is carried out. After the data is ready to be processed, the calculation process is carried out using the C45 and KNN algorithms. after the calculation process is complete. the last process is the process of comparing the results that have been obtained from the calculation.

#### 4.1.1 Collecting Data

Data collection is obtained using questioner form. Data collected using Google Docs and saved as CSV file. After that this file will be imported into the MYSQL database. The data collected from variant menu where is available in the cafe and total data is 1000 data.

**Table 4.1 :** Table Raw Data From Questionare

Jenis kelamin	Usia	Minuman yang sering di pesan ?	Lebih suka kopi dengan rasa apa ?	Makanan yang sering di pesan ?	Rasa makanan apa yang lebih anda sukai dari Ricebowl ?	Merekomendasikan minuman ?	Merekomendasikan makanan ?
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	Ricebowl	Gurih	0	0
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Manual Brewing	Asam	Ricebowl	Gurih	0	1
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Manual Brewing	Asam	Ricebowl	Gurih	1	0
Perempuan	Remaja (Usia 16 - 25 Tahun)	Classic Coffee	Asam	Savoury	Gurih	0	0
Perempuan	Remaja (Usia 16 - 25 Tahun)	Non Coffee	Manis	Pasta	Pedas	0	1
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Squash & Shake	Pait	Sweet	Pedas	1	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	Ricebowl	Pedas	1	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	Ricebowl	Pedas	0	1
Perempuan	Remaja (Usia 16 - 25 Tahun)	Non Coffee	Asam	Ricebowl	Manis	0	0

Perempuan	Remaja (Usia 16 - 25 Tahun)	Chocolate	Manis	Pasta	Manis	1	1
Perempuan	Remaja (Usia 16 - 25 Tahun)	Chocolate	Manis	Sweet	Gurih	1	0
Perempuan	Remaja (Usia 16 - 25 Tahun)	Es kopi susu	Asam	Savoury	Gurih	1	1
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Tea	Manis	Ricebowl	Pedas	0	1
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Tea	Manis	Savoury	Gurih	1	1
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Non Coffee	Pait	Ricebowl	Pedas	0	1
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Es kopi susu	Asam	Sweet	Manis	0	1
Laki - laki	Anak-anak (Usia 8 - 15 Tahun)	Chocolate	Manis	Sweet	Manis	0	0
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Es kopi susu	Asam	Ricebowl	Pedas	0	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Signature	Pait	Ricebowl	Gurih	1	1
Perempuan	Remaja (Usia 16 - 25 Tahun)	Es kopi susu	Asam	Savoury	Pedas	0	1
Laki - laki	Anak-anak (Usia 8 - 15 Tahun)	Chocolate	Manis	Sweet	Manis	1	0
Perempuan	Remaja (Usia 16 - 25 Tahun)	Non Coffee	Manis	Pasta	Pedas	1	0
Perempuan	Remaja (Usia 16 - 25 Tahun)	Classic Coffee	Asam	Savoury	Gurih	1	1
Perempuan	Remaja (Usia 16 - 25 Tahun)	Classic Coffee	Pait	Pasta	Pedas	0	0
Perempuan	Remaja (Usia 16 - 25 Tahun)	Classic Coffee	Pait	Ricebowl	Pedas	0	1
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Classic Coffee	Pait	Ricebowl	Pedas	1	1
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Manual Brewing	Asam	Pasta	Pedas	1	1
Perempuan	Anak-anak (Usia 8 - 15 Tahun)	Chocolate	Manis	Sweet	Manis	0	0
Perempuan	Remaja (Usia 16 - 25 Tahun)	Es kopi susu	Pait	Ricebowl	Pedas	1	0
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Es kopi susu	Pait	Pasta	Pedas	1	1
Perempuan	Dewasa (Usia 26 - 40 Tahun)	Non Coffee	Manis	Savoury	Gurih	1	1
Perempuan	Remaja (Usia 16 - 25 Tahun)	Manual Brewing	Asam	Pasta	Gurih	1	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	Pasta	Pedas	1	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	Pasta	Pedas	0	0
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Manual Brewing	Asam	Ricebowl	Gurih	0	1
Laki - laki	Remaja (Usia 16 - 25 Tahun)	Classic Coffee	Pait	Ricebowl	Manis	1	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Signature	Pait	Ricebowl	Pedas	0	0
Perempuan	Remaja (Usia 16 - 25 Tahun)	Chocolate	Asam	Sweet	Manis	0	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	Savoury	Manis	0	1
Perempuan	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	Sweet	Gurih	0	0
Perempuan	Dewasa (Usia 26 - 40 Tahun)	Squash & Shake	Manis	Savoury	Gurih	1	1
Laki - laki	Anak-anak (Usia 8 - 15 Tahun)	Squash & Shake	Manis	Ricebowl	Manis	1	0
Perempuan	Remaja (Usia 16 - 25 Tahun)	Chocolate	Pait	Sweet	Pedas	0	0

Laki - laki	Remaja (Usia 16 - 25 Tahun)	Signature	Asam	Pasta	Manis	0	1
-------------	-----------------------------	-----------	------	-------	-------	---	---

Table 4.1 the table above is a raw table that comes from the process of filling out the questionnaire that has been done. In this table there are 8 columns which are the initial data which will then be processed. From that date, filtering and sorting will then be carried out to facilitate the calculation process.

## 4.2 Data Pre-processing

After all data imported to database, the next step is to do data preprocessing. In this step data will be filtering and sorting. Filtering is used to filter / display data to be grouped which have predetermined criteria. While sorting is the process of arranging a collection of objects according to certain rules or arrangements.

### 4.2.1 Data Filtering

In this project filtering is used to grouping data according to age range, gender, and separating beetwen food and drink commodities. Filtering process will show data where is needed in accordance with rules has been made to process this data.

**Table 4.2 :** Table Data After Filtering

Jenis kelamin	Usia	Minuman yang sering di pesan ?	Lebih suka kopi dengan rasa apa ?	Merekomendasikan minuman ?
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	1

Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Asam	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Asam	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Asam	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Asam	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Signature	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Signature	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Signature	Pait	1

Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Signature	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Squash & Shake	Manis	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Squash & Shake	Manis	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Tea	Pait	1

Table 4.2 This table is a table that has been filtered based on gender and a certain age range. This filter process is necessary to simplify the data calculation process because there is so much data. With this process the data will be grouped by gender and age range.

#### 4.2.2 Data Sorting

Sorting is process of sorting data that was previously arranged randomly so that it is arranged regularly according to certain rules. In this case the data will sort ascending by name of the menu

Table 4.3 : Table Data After Sorting

Jenis kelamin	Usia	Minuman yang sering di pesan ?	Lebih suka kopi dengan rasa apa ?	Merekomendasikan minuman ?
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	1



Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Asam	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Asam	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Asam	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Manual Brewing	Asam	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Signature	Pait	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Signature	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Signature	Pait	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Signature	Pait	1

Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Squash & Shake	Manis	0
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Squash & Shake	Manis	1
Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Tea	Pait	1

Table 4.3 this table is a table that has been sorted out on the food menu based on alphabetical letters. This process is needed to simplify the data classification process. The filtered and sorted data will then be processed using a predetermined algorithm

### 4.3 Data Processing Using Decision Tree C45

Decision Tree c45 is data mining algorithm which one of the algorithms used to classify or segment or group and is predictive. Classification is one of the processes in data mining that aims to find valuable patterns from relatively large to very large data.

❖ Untuk perhitungan nilai Entropy sbb :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

▪ Keterangan :

- S : himpunan kasus.
- A : fitur.
- n : jumlah partisi S.
- pi : proporsi dari S<sub>i</sub> terhadap S

$$Gain(A) = Entropy(S) - Entropy_A(S)$$

(A)

(B)

**Figure 4.1 :** Decision Tree C45 Formula

Figure 4.1 the picture is a formula for calculating the decision tree c45. in figure 4.1 (a) is the formula for calculating entropy. While in figure 4.1(b) is the last calculation process of the c45 algorithm.

**Table 4.4 :** Result Decision Tree C45

NODE 1.0	JUMLAH TOTAL	JUMLAH YA	JUMLAH TIDAK	ENTROPI	GAIN
TOTAL	28	14	14	1	

MENU MINUMAN						0.30115
	Manual Brewing	8	4	4	1	
	Es kopi susu	3	2	1	0.918296	
	Classic Coffee	10	3	7	0.881291	
	Signature	4	3	1	0.811278	
	Squash & Shake	2	1	1	1	
	Tea	1	1	0	0	
RASA MINUMAN						0.00505
	Asam	7	4	3	0.985228	
	Pait	19	9	10	0.998001	
	Manis	2	1	1	1	

Table 4.4 the table above is an illustration of the calculation process with the c45 algorithm. from the calculation process above, it can be seen that the highest gain is owned by the food menu. while for the taste of food it ranks second

#### 4.4 Data Processing Using K-Nearest Neighbor

The K-Nearest Neighbor (K-NN) algorithm is a classification method for a set of data based on previously classified data learning. Included in supervised learning, where the results of the new query instance are classified based on the majority of the distance proximity of the categories in K-NN.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Figure 4.2 : K-Nearest Neighbor Formula

Figure 4.2 the formula above is a formula for calculating the distance needed in the process of calculating knn. This formula is often called the Euqledian distance. After the results are obtained, the next process is to determine the label of the nearest neighbor.



**Table 4.5 :** Result K-Nearest Neighbor

MENU MINUMAN	RASA MINUMAN	JUMLAH TOTAL	JUMLAH YA	JUMLAH TIDAK	LABEL	DISTANCE	K=1
Non Coffee	Manis	4	2	2	NETRAL	0	NETRAL
Es kopi susu	Asam	2	0	2	TIDAK	2.828427	
Squash & Shake	Manis	2	1	1	NETRAL	2.44949	
Classic Coffee	Pait	4	2	2	NETRAL	0	NETRAL
Manual Brewing	Asam	2	2	0	REKOMENDASI	2.828427	
Tea	Manis	2	0	2	TIDAK	2.828427	
Chocolate	Manis	1	1	0	REKOMENDASI	3.741657	

DATA UJI

MENU MINUMAN	RASA MINUMAN	JUMLAH TOTAL	JUMLAH YA	JUMLAH TIDAK	LABEL
ES KOPI SUSU	MANIS	4	2	2	?

HASIL :

NETRAL

Table 4.5 the table above is a calculation process that has been carried out with the knn algorithm. Test data that does not yet have a recommendation label will be labeled by calculating the distance from its nearest neighbor using the Euclidian distance formula. After the calculation process is carried out, it is found that the label that matches the nearest neighbor is not a recommendation

## 4.5 Data Processing Using Confusion Matrix

Confusion Matrix is a performance measurement for machine learning classification problems where the output can be two or more classes. The Confusion Matrix is a table with 4 different combinations of predicted values and actual values. There are four terms that represent the results of the classification process in the confusion matrix, namely True Positive, True Negative, False Positive, and False Negative.

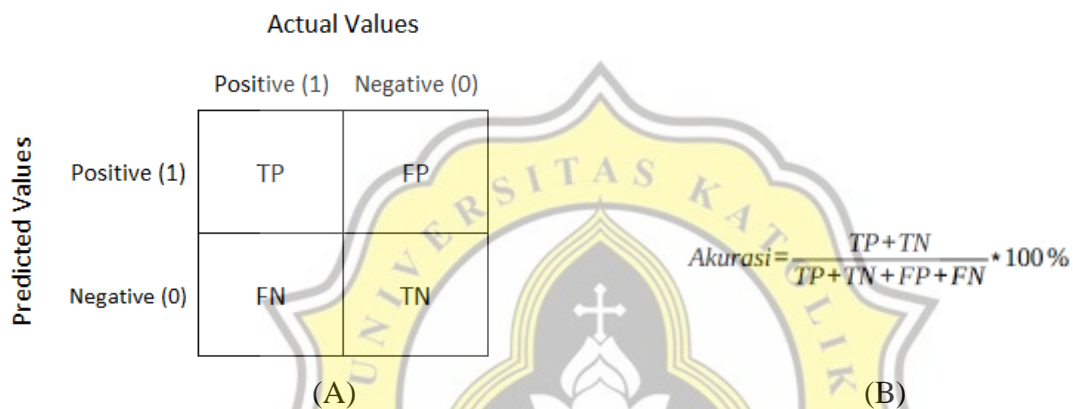


Figure 4.3 : Confusion Matrix

Figure 4.3 Confusion Matrix is a performance measurement for machine learning classification problems where the output can be two or more classes. In figure 4.3 (a) show that Confusion Matrix had a table with 4 different combinations of predicted values and actual values. While in figure 4.3 (b) is a formula for calculating accuracy with parameters that have been obtained in the previous process.

## 4.6 Design

### 4.6.1. Flow Chart

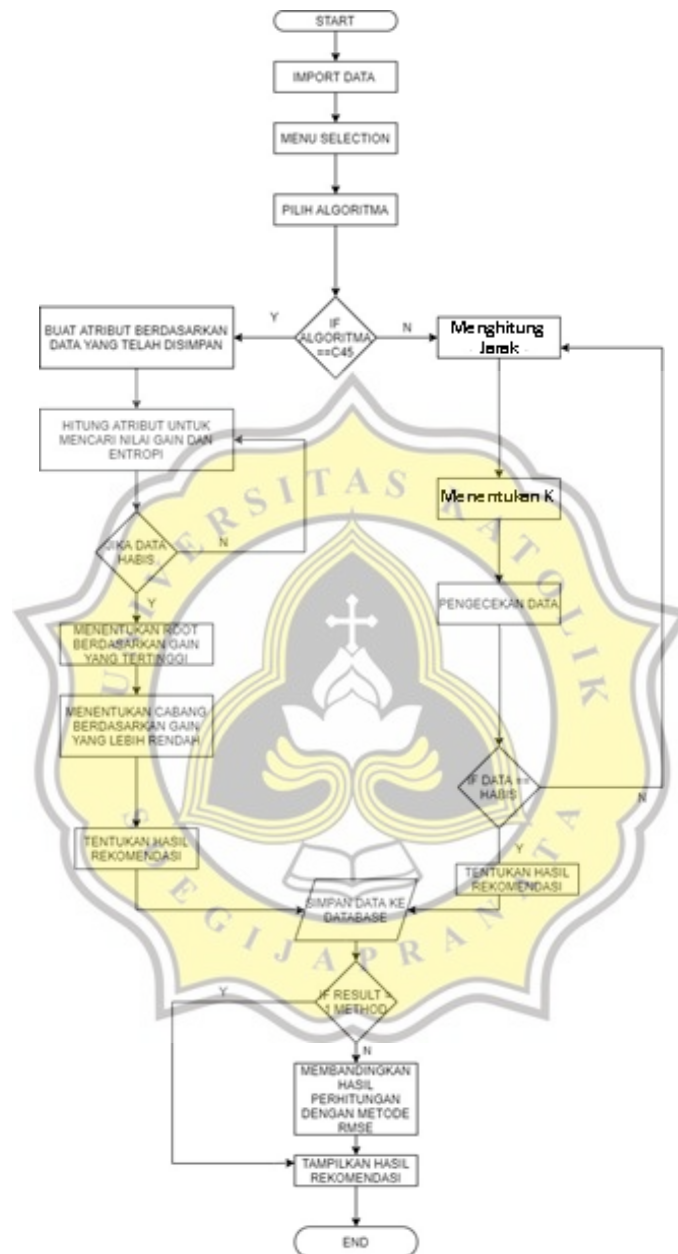


Figure 4.4 : Flowchart

Figure 4.4 the first step to do is import data to database. After it done, user must chose variable and algorithm and submit process. Preprocessing data will do automatically since user selected variable. The result from this project will automatically save to database. Each result of algorithm will be compared in the end classification process.

